1. (4 puntos) Sean $X_1, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas con función de densidad para cada X_i

$$f(x;\theta) = \frac{x^2}{2\theta^3} e^{(-x/\theta)}, \quad x > 0$$

donde $\theta > 0$ es un parámetro. Sabemos que $E(X) = 3\theta$ y que $Var(X) = 3\theta^2$.

   (a) Hallar el estimador por el método de los momentos ($\hat{\theta}_{MOM}$) del parámetro θ.
   (b) Calcular la función de verosimilitud $\ell(\theta)$ para esta muestra aleatoria.
   (c) Hallar el estimador máximo verosímil ($\hat{\theta}_{MLE}$) de θ.
   (d) Demostrar que ambos estimadores en (a) y (c) son insesgados.
   (e) Demostrar que $\sum_{i=1}^{n} X_i$ es un estadístico suficiente para el parámetro θ.
   (f) Hallar la cota de Cramer-Rao para la varianza de un estimador insesgado de θ.
   (g) ¿Cuál es el valor de la información de Fisher $\mathcal{I}_X(\theta)$ en una observación individual de esta densidad?
   (h) ¿Es ($\hat{\theta}_{MLE}$) el estimador insesgado de varianza mínima de θ?
   (i) Hallar el estadístico del test de la razón de verosimilitud para contrastar la hipótesis $H_0 : \theta = \theta_0$ contra la alternativa $H_1 : \theta \neq \theta_0$.
   (j) Supongamos que una muestra aleatoria de tamaño n=27 produce $\sum_{i=1}^{27} x_i = 108$. Utilizando el contraste construido en (i) ¿podemos rechazar la hipótesis $H_0 : \theta = 1$ frente a la alternativa $H_1 : \theta \neq 1$ a un nivel de significación de $\alpha = 0.10$?

2. (2 puntos) Los tiempos de vida de unos ventiladores diésel (X) se pueden considerar distribuidos como una distribución Gamma ($\alpha; \beta$) con α conocida y β desconocido.

   Se toma una muestra aleatoria $X_1, \ldots, X_n$ de dichos tiempos de vida. Sabiendo que:

$$\frac{2\sum_{i=1}^{n} X_i}{\beta} \sim \chi^2_{4n}$$

   (a) Construir un intervalo de confianza con nivel de confianza $(1-\alpha)\%$ para el parámetro desconocido β.
   (b) Dada una muestra aleatoria de n=15 datos con $\bar{X} = 135$ (en miles de horas) construir un intervalo de confianza al 95 % para β.
   (c) Los ingenieros constructores de los ventiladores creen que el parámetro β (tasa de fallos) debería estar por encima de 150. En vista a los resultados obtenidos en (b) ¿es esto cierto?

Nota: Responder cada pregunta en una hoja de respuestas diferente.

3. (2 puntos) En el siglo XIX, un físico escocés, James D. Forbes, quiso determinar la altitud sobre el nivel del mar a través de la medición del punto de ebullición del agua. Él sabía que la altitud podía determinarse a través de la presión atmosférica, ya que a mayor altitud, menor presión. Hooker realizó en paralelo un experimento similar. Así, Forbes y Hooker midieron la presión atmosférica y la temperatura de ebullición del agua en 17 y 31 puntos respectivamente. Los datos con los que se han construido los modelos se han sacado de Weisberg (1985).

MODELO FORBES

```
lm(formula = logP ~ logT)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0077792 -0.0032640 -0.0001208  0.0000928  0.0297415

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.10253    0.39161  -48.78   <2e-16 ***
logT          4.20078    0.07371   56.99   <2e-16 ***MO
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.008339 on 15 degrees of freedom
Multiple R-squared:  (*),       Adjusted R-squared:  ---
F-statistic:  3248 on 1 and 15 DF,  p-value: < 2.2e-16
```

MODELO HOOKER

```
lm(formula = logPhook ~ logThook)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0162907 -0.0049490  0.0000369  0.0065430  0.0178281

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -18.32959    0.17906  -102.4   <2e-16 ***
logThook      4.05488    0.03407   119.0   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.008394 on 29 degrees of freedom
Multiple R-squared:  0.998,     Adjusted R-squared:  0.9979
F-statistic: 1.416e+04 on 1 and 29 DF,  p-value: < 2.2e-16
```

(a) Calcular el valor del coeficiente de determinación (*) para el modelo de Forbes teniendo en cuenta que la varianza muestral (S^2) del logaritmo de las 17 presiones medidas es 0.01418

(b) Haz las pruebas inferenciales que creas convenientes para ver si los modelos obtenidos son iguales o no.

(c) Para realizar pruebas inferenciales con los modelos especificados, ¿qué condiciones se deben cumplir y cómo se validan?

Nota: Responder cada pregunta en una hoja de respuestas diferente.

4. (2 puntos) Un artículo publicado en la Revista de la Sociedad Electroquímica (Vol. 139, No. 2, 1992, pp. 524-532) describe un experimento para investigar la deposición de vapor de baja presión de polisilicio. El experimento se llevó a cabo en un reactor de gran capacidad en Sematech en Austin, Texas. El reactor tiene varias posiciones de la oblea, y **cuatro de estas posiciones se seleccionan al azar**. La variable de respuesta es la uniformidad de espesor de la película. Se realizaron tres repeticiones para cada posición. El promedio de las 12 observaciones fue $\bar{X} = 2.33$.

   La tabla ANOVA es la siguiente:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Posiciones de la oblea		16.2198			0.008
Residuals		5.2175			

   (a) Indica las fuentes de variabilidad que intervienen en el experimento y la ecuación del modelo para analizar los datos.

   (b) Rellena los espacios vacíos de la tabla ANOVA.

   (c) Indica las hipótesis nula y alternativa del contraste. ¿Cuál es la decisión que tomaremos basada en la tabla anterior? Considerar $\alpha = 0.05$.

   (d) Estimar todos los parámetros del modelo.

   (e) ¿Cuales son los cambios en el modelo, hipótesis o en la interpretación de los resultados si las 4 posiciones de la oblea son las únicas que pueden usarse. Indicar las conclusiones del estudio en esta situación.

Nota: Responder cada pregunta en una hoja de respuestas diferente.

0.1 Soluciones

1. Problema 1

   (a) $E(X) = 3\theta \Rightarrow \hat{\theta}_{MOM} = \bar{X}/3$

   (b) $\ell(\theta) = \frac{\prod_{i=1}^n x_i^2}{2^n \theta^{3n}} e^{-\sum_{i=1}^n x_i/\theta}$

   (c) $\frac{\partial \log(\ell(\theta))}{\partial \theta} = \frac{-3n}{\theta} + \frac{\sum x_i}{\theta^2} = 0 \Rightarrow \hat{\theta}_{MLE} = \bar{X}/3$

   (d) Puesto que ambos son iguales basta con comprobarlo en un caso $E(\bar{X}/3) = 3\theta/3 = \theta$.

   (e) Si $T = \sum_{i=1}^n x_i$, entonces

   $$f(x_1, \ldots, x_n; \theta) = \frac{e^{-T/\theta}}{2^n \theta^{3n}} \prod_{i=1}^n x_i^2 = g(T, \theta) \cdot h(x_1, \ldots, x_n)$$

   (f)

   $$\frac{\partial \ln f}{\partial \theta} = -3/\theta + x/\theta^2$$
   $$\frac{\partial^2 \ln f}{\partial \theta^2} = 3/\theta^2 - 2x/\theta^3$$
   $$E(3/\theta^2 - 2x/\theta^3) = -3/\theta^2$$
   $$CRB(\theta) = \frac{1}{\frac{3n}{\theta^2}} = \frac{\theta^2}{3n}$$

   (g) $\mathcal{I}_X(\theta) = 3/\theta^2$

   (h) $Var(\bar{X}/3) = \frac{3\theta^2}{9n} = \frac{\theta^2}{3n} = CRB(\theta)$

   (i) $\Lambda = \frac{\frac{\exp(-\sum_{i=1}^n x_i/\theta_0)}{\theta_0^{3n}}}{\frac{\exp(-3n)}{\hat{\theta}^{3n}}} = \left(\frac{\hat{\theta}}{\theta_0}\right)^{3n} e^{3n - \frac{\sum x_i}{\theta_0}}$

   (j) $\hat{\theta} = 108/81$.

   $$\Lambda = (108/81)^{81} e^{81-108} = 0.024779$$
   $$U = -2 \ln \Lambda = 7.3955$$
   $$\chi_{1,0.1} = 2.70554 \Rightarrow \text{ Rechazamos } H_0$$

2. Problema 2

   (a) Sabiendo que

   $$\frac{2\sum_{i=1}^n X_i}{\beta} \sim \chi^2_{4n}$$

   tenemos que $\frac{2\sum_{i=1}^n X_i}{\beta}$ es un estadístico pivotal, puesto que su distribución no depende del parámetro desconocido β. Por tanto el IC $1-\alpha$ será:

   $$P(\chi^2_{1-\alpha/2} \leq \frac{2\sum_{i=1}^n X_i}{\beta} \leq \chi^2_{\alpha/2}) = 1 - \alpha$$
   $$P(\frac{2\sum_{i=1}^n X_i}{\chi^2_{1-\alpha/2}} \leq \beta \leq \frac{2\sum_{i=1}^n X_i}{\chi^2_{\alpha/2}}) = 1 - \alpha$$

   Así, $\frac{2\sum_{i=1}^n X_i}{\chi^2_{1-\alpha/2,4n}}; \frac{2\sum_{i=1}^n X_i}{\chi^2_{\alpha/2,4n}}$ es dicho intervalo.

Nota: Responder cada pregunta en una hoja de respuestas diferente.

(b) En el caso particular el intervalo obtenido es $(48.62; 100.04)$.

(c) Claramente 150 no está incluido en el intevalo, por lo que no es creible la afirmación.

3. Problema 3

(a)
$$SQT = (n-1)S_y^2 = 16 \cdot 0.01418 = 0.22688$$
$$SQR = (n-p)S_r^2 = 15 \cdot 0.008339^2 = 0.001043084$$
$$R^2 = 1 - (0.001043084/0.22688) = 0.99538$$

(b) Se pueden hacer muchas pruebas, pero quizá lo más sencillo sería comprobar que se solapan los IC para los coeficientes.
Modelo Forbes: IC para $\beta_0 : -19.10 \pm 2 \cdot 0.3916 = (-19.88; -18.316)$
IC para β_1: $4.2 \pm 2 \cdot 0.0737 = (4.0526; 4.347)$.
Modelo Hooker: IC para $\beta_0 : -18.33 \pm 2 \cdot 0.17906 = (-18.687; -17.97)$
IC para β_1: $4.05 \pm 2 \cdot 0.03407 = (3.986; 4.12)$.
Como los intervalos encontrados se solapan, no podemos decir que los modelos sean distintos.

(c) Se tienen que validar las hipòtesis del modelo teórico: Linealidad, variancia constante, normalidad e independencia. Se hace mediante el análisis de los residuos.

4. Problema 4

(a) La fuente de variabilidad controlada es la posición de la oblea a la que hay que añadir la variación individual que conforma el error.
La ecuación del modelo es
$$y_{ij} = \mu + A_i + \epsilon_{ij}$$
donde A_i representa la posición de la oblea, factor aleatorio con i=1,2,3,4 y donde ϵ_{ij} es el término de error con j=1,2,3.
En este modelo $A_i \sim N(0, \sigma_A)$ y $\epsilon_{ij} \sim N(0, \sigma_\epsilon)$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Posiciones de la oblea	3	16.2198	5.4066	8.29	0.008
Residuals	8	5.2175	0.65218		

(b)

(c) $H_0 : \sigma_A = 0 \ vs. \ H_1 : \sigma_A \neq 0$. Rechazamos la hipótesis nula y aceptamos que existe una variabilidad aportada por la posición de la oblea.

(d) $\hat{\mu} = 2.33, \hat{\sigma_\epsilon^2} = 0.65218, \hat{\sigma_A^2} = (5.4066 - 0.65218)/3 = 1.5848$.

(e) Pasa a ser un modelo de efectos fijos donde la hipótesis nula es que no hay diferencias entre las cuatro posiciones
$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$
$H_0 : \alpha_i = 0 \forall i \ vs. \ H_1 : \exists \alpha_i \neq 0$. Igualmente rechazamos la hipótesis nula y concluimos que las cuatro posiciones no son iguales.

Nota: Responder cada pregunta en una hoja de respuestas diferente.